Quantitative positional proteomics for the identification of CLAP1 protease substrates

Stefan Niedermaier[1], Daniela Pedroso[2], Pitter F. Huesgen[1], Isaura Simões[2]

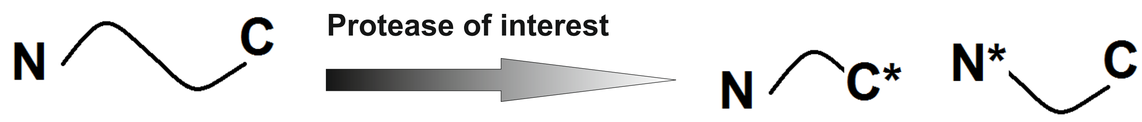
[1] *Forschungszentrum Jülich, Central Institute for Engineering, Electronics and Analytics (ZEA-3), Germany*
[2] *Centre for Neuroscience and Cell Biology, University of Coimbra & Biocant-Biotechnology Innovation Centre, Coimbra, Portugal*

Background

- Proteolysis is an essential, irreversible posttranslational protein modification, involved in the regulation of many cellular processes.
- Over 800 different proteases are encoded by the *Arabidopsis thaliana* genome, illustrating the importance of these enzymes in higher plants. However, for most of these proteases no *in vivo* substrates have been identified yet and their physiological functions still remain unclear.
- The amino acid sequence of CLAP1 shows some atypical features such as its unusual high amount of cysteine residues that might indicate a regulation by stress-induced redox changes in the cell and a predicted putative DNA-binding sequence, implying some yet unknown DNA interaction potentially involved in the localization or activity regulation of the protease. However, no physiological substrates have been described for CLAP1 yet and no CLAP1 knock-out mutant is available that could provide a clue about the physiological function of this protease.

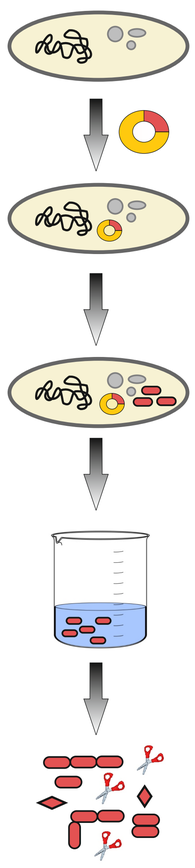
Degradomics: Proteome-wide identification of protease substrates

- Cleavage a of substrate by a protease of interest generates new N and C termi not present in the control samples



- **Challenge:** relatively small number of terminal peptides is lost in the majority of internal tryptic peptides → high sample complexity
- **„TAILS“ approach:** Label and enrich terminal peptides (including protease generated N termini) for identification

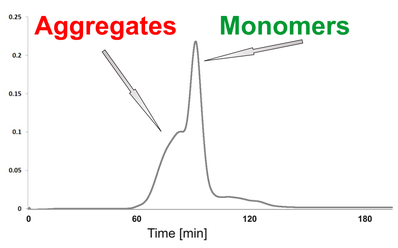
Recombinant expression and purification of CLAP1

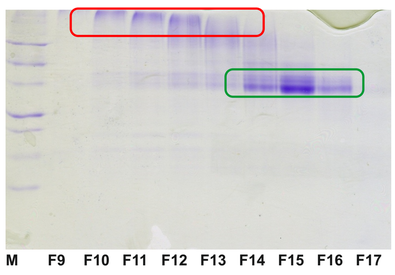


1. Transformation of *E. coli* with recombinant vectors containing the CLAP1 wildtype and active site mutant sequence
2. Expression of CLAP1 as inclusion bodies
3. Extraction of CLAP1 inclusion bodies and refolding by dialysis

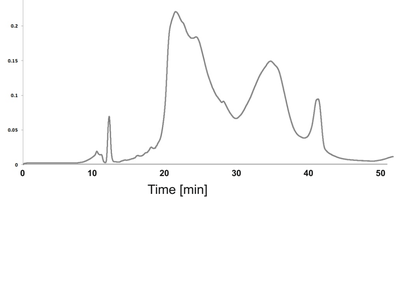
Refolding yields correctly folded CLAP1 as well as misfolded, denatured and aggregated CLAP1.

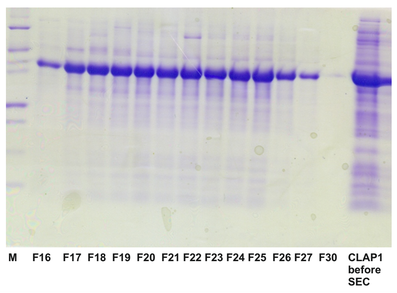
4. In the first purification step, aggregates are separated from monomers by size exclusion chromatography (SEC)

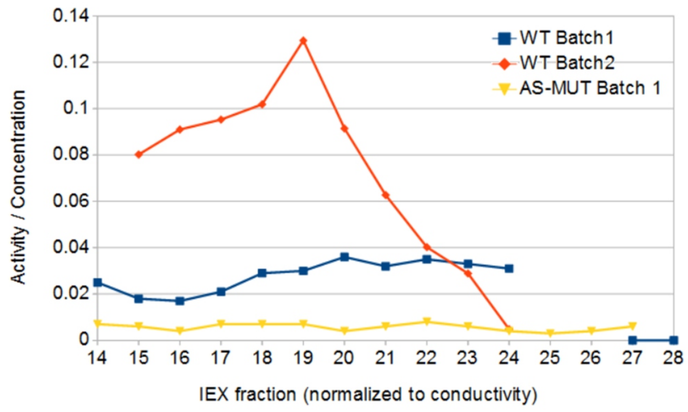




5. Differently folded CLAP1 forms are seperated by anion exchange chromatography due to their different surface charge

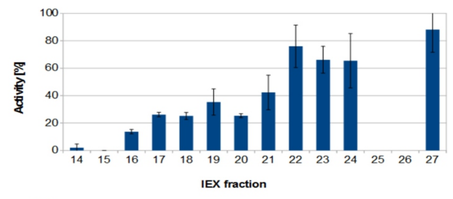






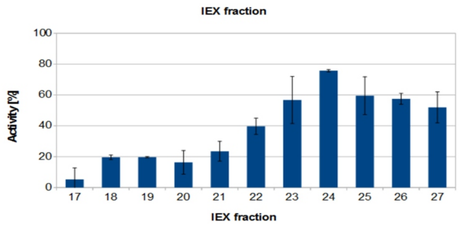
6. We expressed and purified two independant batches of CLAP1 under slightly different refolding conditions. In both experiments a significant protease activity against a fluorogenic protease substrate (MCA-Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys-DNP) was observed in all relevant IEX fractions.The active site mutant showed no significant protease activity in the corresponding IEX fractions.

Batch 1:
0.25 mg / mL
72 h refolding



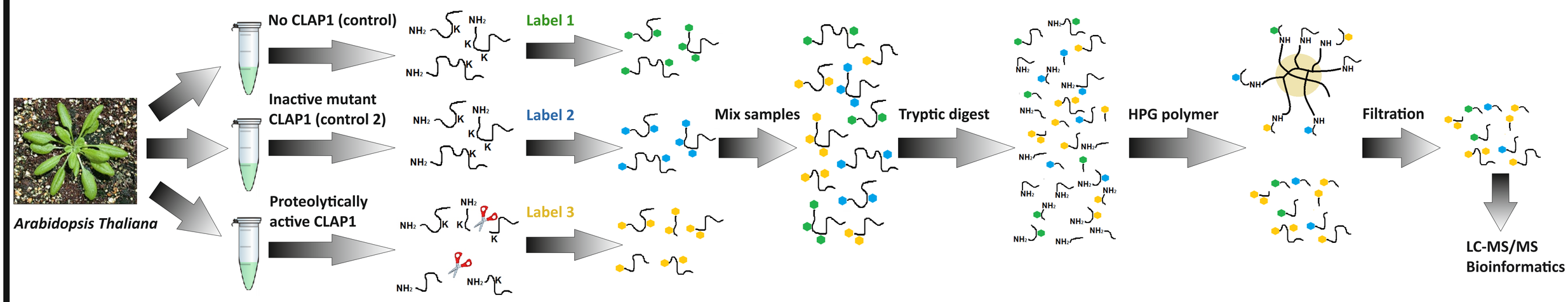
Batch 2:
0.50 mg / mL
144 h refolding

c (Pepstatin) = 1µM
n = 2



7. We observed that the fractions that elute early from IEX are much more sensitive to inhibition by the specific aspartic protease inhibitor pepstatin compared to the later eluting fractions. This indicates that IEX leads to a separation of differently folded CLAP1 forms with different activity / inhibition profiles.

„N-TAILS“ for CLAP1 substrate identification



Prepare proein extract (e.g.total protein, chloroplast or membrane extract)

Protein extracts are incubated with recombinant wildtype and active-site mutant CLAP1. Proteolytic activity generates protein N termini that are not present in the control samples.

Primary amines of N termini and lysines are labeled and blocked (e.g. stable isotope dimethyl labeling or TMT labeling).

Samples are mixed and digested with trypsin. The labeled N termini of the proteins remain blocked, while the newly generated internal tryptic peptides carry unblocked N termini with free primary amines.

A hyperbranched polyglycerol and aldehyde (HPG) polymer is added to the sample to react with the primary amines at the free N termini of tryptic peptides. The blocked original protein N termini do not bind to the polymer and can be separated by filtration.

Evaluation of „N-TAILS“ experiments

Evaluation of TAILS LC-MS/MS data requires at least 2 Max-Quant database searches and some considerations:

- **Digestion mode:** Semispecific free N terminus
- **1st search:** Quantification set to labeled N termini and labeled lysines
  -> Identify original and protease generated protein N termini
- **2nd search:** Quantification set to labeled lysines only. Variable modifications (acetylation, Glu-> pyro Glu, Gln-> pyro Glu) on N termini
  -> Identify N-terminal peptide modifications that are also enriched by „N-TAILS“

Additional considerations when using stable isotope dimethyl labeling:

- **3rd search:** No quantification, variable monomethylation on N termini and lysines
  -> Control of labeling efficiency
- Select Arg-C (missed cleavage > 2) as digestion protease since trypsin proteolysis at dimethylated lysins is very rare

Conclusion and Outlook

- Wildtype CLAP1 and a CLAP1 active-site mutant were successfully expressed
- Recombinant CLAP1 wildtype is active at acidic pH (~4.0) and showed sensitivity to the specific aspartic protease inhibitor pepstatin
- No proteolytic activity of CLAP1 active site mutant (control sample)
- Proteomics-based identification of CLAP1 substrates on relevant *Arabidopsis thaliana* extracts is in preparation

Acknowledgement:

This project is supported by a PPP grant of the German Academic Exchange Service with funding from the Federal Ministry of Education and Research (Project-ID 57128819 to PFH) and the Fundação para a Ciência e a Tecnologia (FCT) (grant: Scientific and Technological Bilateral Aggreement 2015/2016 to IS).